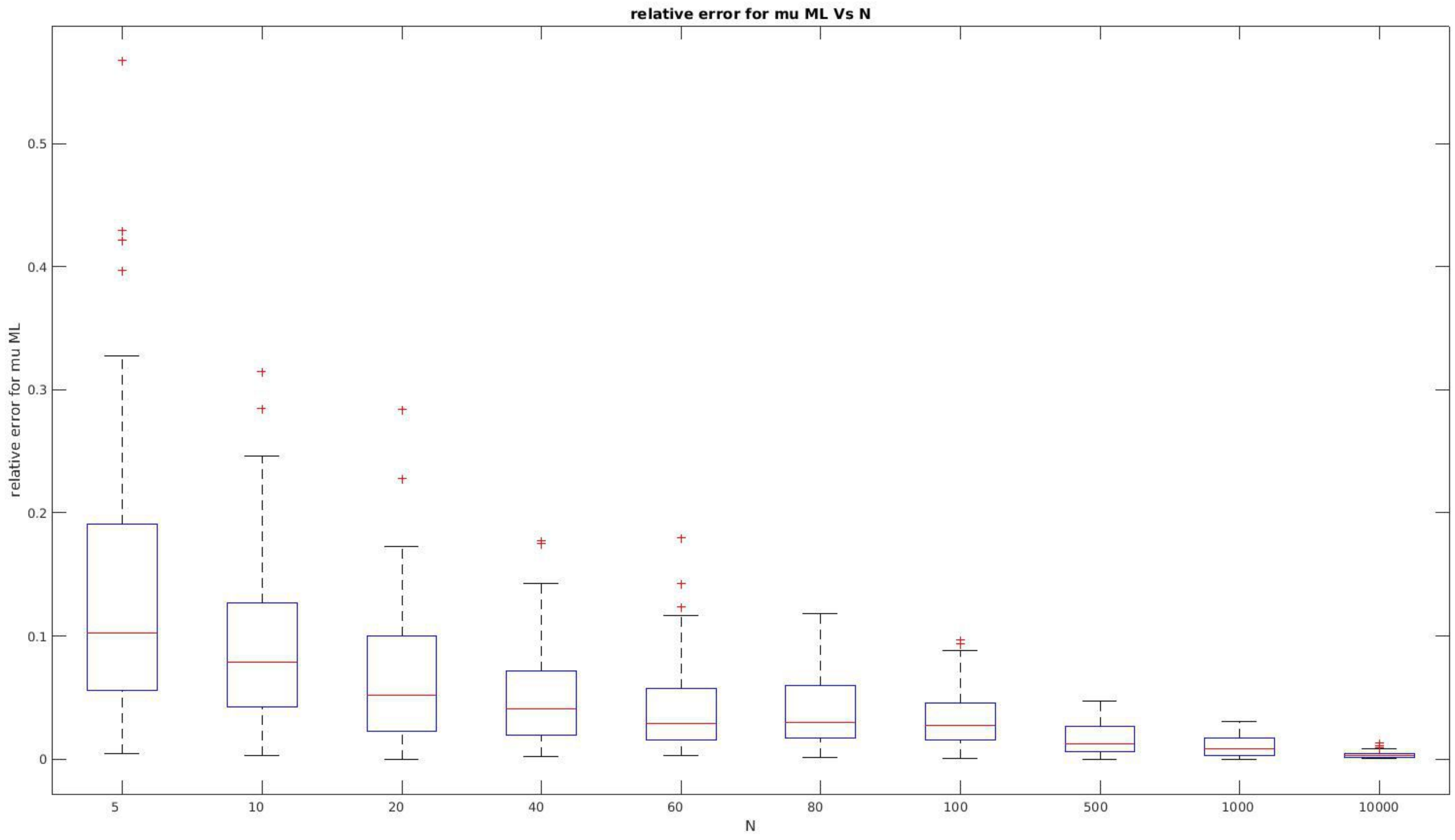
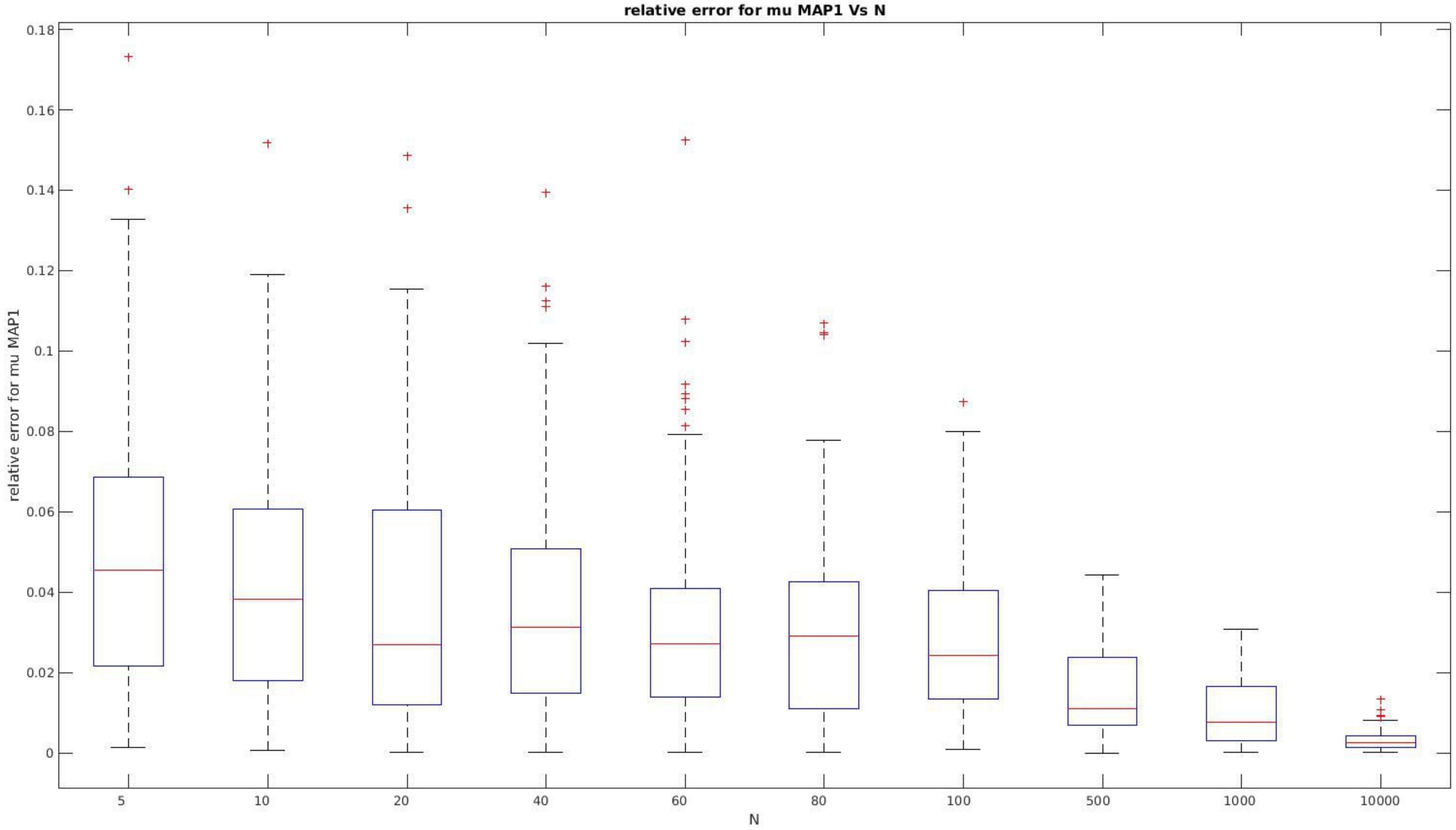
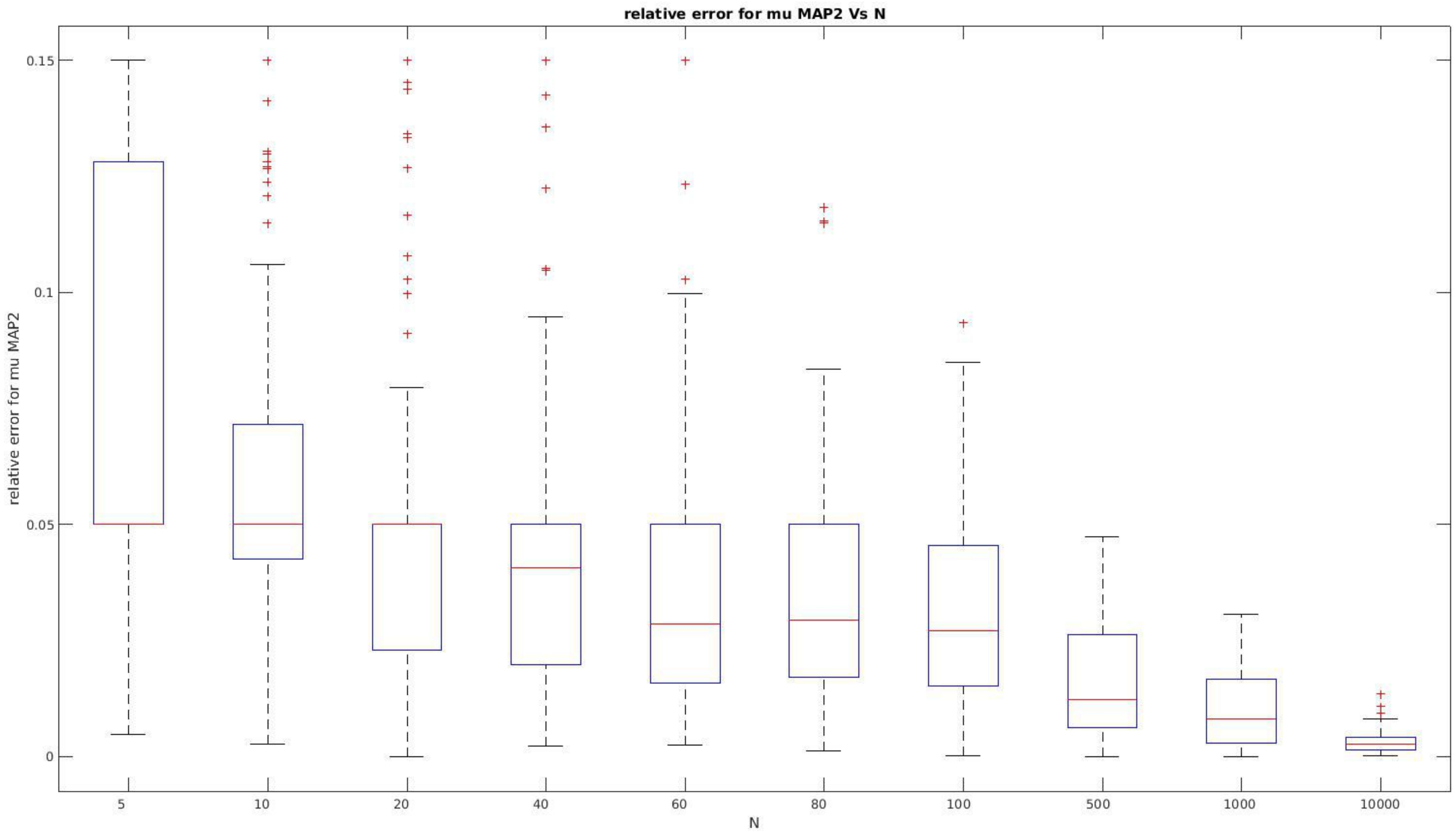


relative error for mu ML Vs N
relative error for mu ML
0
0.1
0.2
0.3
0.4
0.5
5
10
20
40
60
80
100
500
1000
10000
N


relative error for mu MAP1 Vs N
relative error for mu MAP1
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
5
10
20
40
60
80
100
500
1000
10000
N


relative error for mu MAP2 Vs N
relative error for mu MAP2
0.15
0.1
0.05
0
5
10
20
40
60
80
100
500
1000
10000
N

Problem - 1 :-

consider $x_1, x_2, x_3 \dots x_N$ are the N points drawn from the Gaussian distribution. All of them are i.i.d random Variables.

→ $\hat{\mu}^{ML}$:

$$\text{Likelihood} = P(x_1, x_2 \dots x_N \mid \mu) \propto \prod_{i=1}^{N} e^{-\left(\frac{(x_i-\mu)^2}{2\sigma^2_{true}}\right)}$$

So, we need to maximize $P(x_1, x_2 \dots x_N \mid \mu)$

$\Rightarrow$ maximize $\log P(x_1, x_2 \dots x_N \mid \mu)$

$\Rightarrow$ maximize $\sum_{i=1}^{N} -\left(\frac{(x_i-\mu)^2}{2\sigma^2}\right)$

$\Rightarrow$ Minimize $\sum_{i=1}^{N} (x_i-\mu)^2$

differentiate with respect to μ and equal to zero

$$\Rightarrow \sum_{i=1}^{N} x_i - n\mu = 0 \Rightarrow \mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

$$\therefore \boxed{\hat{\mu}^{ML} = \frac{\sum_{i=1}^{N} x_i}{N}}$$

→ $\hat{\mu}^{MAP}$:

$$\text{prior} = G(\mu; \mu_{prior}, \sigma^2_{prior})$$

$$\mu_{prior} = 10.5, \quad \sigma^2_{prior} = 1$$

$$\text{Posterior} = P(\mu \mid x_1, x_2 \dots x_N) = \frac{P(x_1, x_2 \dots x_N \mid \mu)\, P(\mu)}{P(x_1 \dots x_N)}$$

$$G(z; \mu_1, \sigma_1^2) \cdot G(z; \mu_2, \sigma_2^2) = G(z; \mu_3, \sigma_3^2);$$

$$\text{where } \mu_3 = \frac{\mu_2\sigma_1^2 + \mu_1\sigma_2^2}{\sigma_1^2+\sigma_2^2}, \quad \sigma_3^2 = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}$$

Negative exponent in $P(x_1, x_2 \dots x_N \mid \mu)$ can be written as

$$\left(N\mu^2 - 2\left(\Sigma x_i\right)\mu\right) / 2\sigma^2_{true} + c \rightarrow \text{independent of } \mu$$

$$= \left(\mu^2 - 2\left(\frac{\Sigma x_i}{N}\right)\mu\right) \Big/ \frac{2\sigma^2_{true}}{N} + c$$

$$\Rightarrow P(x_1, x_2, \dots x_N \mid \mu) \propto G\left(\mu; \frac{\Sigma x_i}{N}, \frac{\sigma^2_{true}}{N}\right)$$

Hence posterior $\propto G\left(\mu;\ \frac{\Sigma x_i}{N},\ \frac{\sigma^2_{true}}{N}\right) G\left(\mu;\ \mu_{prior},\ \sigma^2_{prior}\right)$

$\propto G(\mu;\ \mu^*,\ \sigma^{2*})$

where
$$\mu^* = \frac{\left(\Sigma x_i/N\right)\sigma^2_{prior} + \left(\sigma^2_{true}/N\right)\mu_{prior}}{\sigma^2_{prior} + \sigma^2_{true}/N}$$

$$\sigma^{2*} = \frac{\sigma^2_{prior} \cdot \sigma^2_{true}/N}{\sigma^2_{prior} + \sigma^2_{true}/N}$$

$\hat{\mu}^{MAP1} = \mu^*$ (Gaussian PDF is maximum at mean)

$$\therefore \boxed{\hat{\mu}^{MAP1} = \frac{\left(\Sigma x_i/N\right) + \left(\frac{16}{N}\right)10.5}{1 + \frac{16}{N}}} \rightarrow \mu_{prior} = 10.5,\ \sigma_{true} = 4,\ \sigma_{prior} = 1$$

$$\therefore \hat{\mu}^{MAP1} = \frac{\Sigma x_i + 168}{N + 16}$$

→ $\hat{\mu}^{MAP2}$:- prior :- uniform distribution over (9.5 11.5)

$\Rightarrow P(\mu) = \frac{1}{2}$ $\mu \in (9.5\ 11.5)$

$= 0$ else where

$$\text{posterior} = P(\mu | x_1, x_2 .. x_N) = \frac{P(x_1, x_2 .. x_N | \mu)\ P(\mu)}{P(x_1, .. x_N)}$$

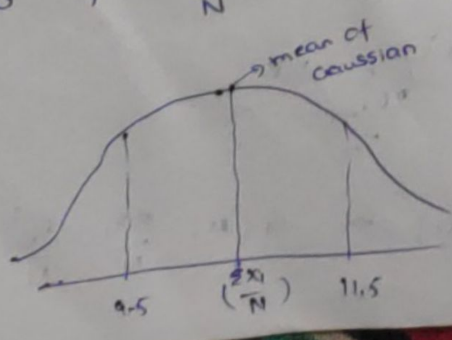
$P(x_1, x_2 .. x_N) \rightarrow$ independent of μ

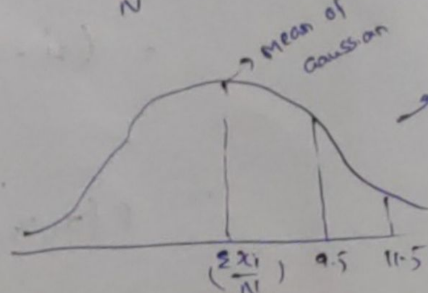
posterior $\propto \frac{1}{2} \prod_{i=1}^{N} e^{-(x_i - \mu)^2 / 2\sigma^2_{true}}$ $\mu \in (9.5\ 11.5)$

$\prod_{i=1}^{N} e^{-\frac{(x_i - \mu)^2}{2\sigma^2_{true}}}$ is maximum when $\mu = \frac{\Sigma x_i}{N}$ (done in previous part)

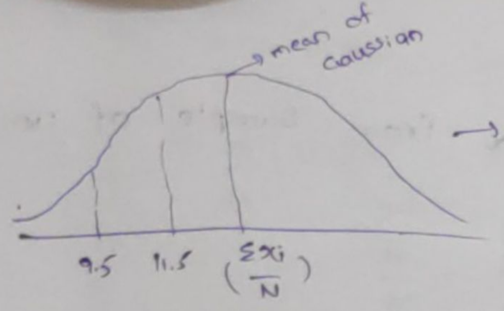
So if $\frac{\Sigma x_i}{N} \in [9.5\ 11.5]$, $\hat{\mu}^{MAP2} = \frac{\Sigma x_i}{N}$ (case-i)

case-(i)





case-(ii)



→ Case - (iii)

In case (ii), $\frac{\sum x_i}{N} < 9.5$, $\hat{\mu}^{MAP2} = 9.5$ because for this value we can maximize posterior (from graph drawn)

Similarly in case-(iii), where $\frac{\sum x_i}{N} > 11.5$, $\hat{\mu}^{MAP2} = 11.5$

$$\Rightarrow \hat{\mu}^{MAP2} = \frac{\sum x_i}{N} \quad \text{if} \quad \frac{\sum x_i}{N} \in [9.5 \;\; 11.5]$$

$$= 9.5 \quad \text{if} \quad \frac{\sum x_i}{N} < 9.5$$

$$= 11.5 \quad \text{if} \quad \frac{\sum x_i}{N} > 11.5$$

* clearly as 'N' increases the error decreases, the box-plot tends closer to zero. (from box-plot graphs)

* Among the three estimates $\hat{\mu}^{MAP1}$ is the best one as error will be minimum in this case. (from box plot graphs)

Problem - 2 :-

Let $x_1, x_2 \ldots x_N$ is resultant data from sample of uniform distribution on $(0,1)$

$$p(x) = \frac{1}{1-0} = 1$$

→ we have to transform resulting data x to y

$$y = \frac{-1}{\lambda} \log(x) \quad \text{where } \lambda \text{ is some parameter} \quad \to ①$$

Let $y = g(x) \Rightarrow x = g^{-1}(y)$

$$\therefore p(y) = p(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| \quad \left(\text{as } \underline{p(y)dy = p(x)dx}\right)$$

$$p(g^{-1}(y)) = p(x) = 1 \quad (\text{from above})$$

from eq ①

$$\log x = -\lambda y$$

$$\Rightarrow x = e^{-\lambda y} = g^{-1}(y)$$

$$\Rightarrow \frac{d}{dy} g^{-1}(y) = \frac{d}{dy}(x)$$

$$= \frac{d}{dy}(e^{-\lambda y}) = -\lambda e^{-\lambda y}$$

$$\therefore p(y) = p(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|$$

$$= 1 \times |-\lambda . e^{-\lambda y}|$$

Analytic form :- $\underline{p(y) = |\lambda| e^{-\lambda y}}$

→ $\hat{\lambda}^{ML}$:- likelihood $= p(y_1, y_2 \ldots y_N | \lambda) = \prod_{i=1}^{N} |\lambda| e^{-\lambda y_i} \to ①$

$$\text{likelihood} = |\lambda|^N e^{-\lambda \sum_{i=1}^{N} y_i}$$

λ = is positive

$$\Rightarrow \text{likelihood} = \lambda^N e^{-\lambda \sum_{i=1}^{N} y_i}$$

$$\Rightarrow \text{maximize} \quad \lambda^N e^{-\lambda \sum_{i=1}^{N} y_i}$$

$\Rightarrow$ maximize $\log \lambda^N e^{-\lambda \sum_{i=1}^{N} y_i}$ $\rightarrow$ differentiate w.r.t 'λ'

↳ equate to '0'

$$\Rightarrow \frac{d}{d\lambda}\left(N \log^{\lambda} - \lambda\left(\sum_{i=1}^{N} y_i\right)\right) = 0$$

$$\Rightarrow \frac{N}{\lambda} = \sum_{i=1}^{N} y_i$$

$$\Rightarrow \boxed{\hat{\lambda}^{ML} = \frac{N}{\sum_{i=1}^{N} y_i}}$$

$\rightarrow$ $\hat{\lambda}^{\text{posterior Mean}}$:-

Prior = Gamma prior where $\alpha = 5.5$, $\beta = 1$

$$\gamma(\lambda|_{\alpha,\beta}) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}$$

$$\text{Posterior} = P(\lambda|_{y_1, y_2, \dots y_N}) = \frac{P(y_1, y_2 \dots y_N | \lambda) \cdot P(\lambda)}{P(y_1, y_2 \dots y_N)}$$

$$\Rightarrow \text{posterior} = \frac{\lambda^{n} e^{-\lambda \Sigma y_i} \lambda^{\alpha-1} e^{-\beta\lambda}}{\int_{\lambda} \lambda^{n} e^{-\lambda \Sigma y_i} \lambda^{\alpha-1} e^{-\beta\lambda} \, d\lambda}$$

$$\text{Denominator} = \int_0^{\infty} \lambda^{n} e^{-\lambda \Sigma y_i} \lambda^{\alpha-1} e^{-\beta\lambda} \, d\lambda \quad (\lambda > 0)$$

$$= \int_0^{\infty} \lambda^{n+\alpha-1} e^{-\lambda(\beta + \Sigma y_i)} \, d\lambda$$

$$\text{Denominator} = \frac{(n+\alpha-1)!}{(\beta + \Sigma y_i)^{n+\alpha}} \quad \left(\text{as } \int_0^{\infty} x^{n} e^{-ax} dx = \frac{n!}{a^{n+1}} \text{ for n} \right.$$

$$\Rightarrow \text{posterior} = \frac{\lambda^{n+\alpha-1} e^{-\lambda(\beta + \Sigma y_i)}}{\frac{(n+\alpha-1)!}{(\beta + \Sigma y_i)^{n+\alpha}}}$$

$$\underset{P(\lambda|_{\text{data}})}{E} [\lambda] = \text{posterior Mean}$$

$$= \int_0^{\infty} P(\lambda|_{\text{data}}) \, \lambda \, d\lambda \quad (\text{as } \lambda > 0)$$

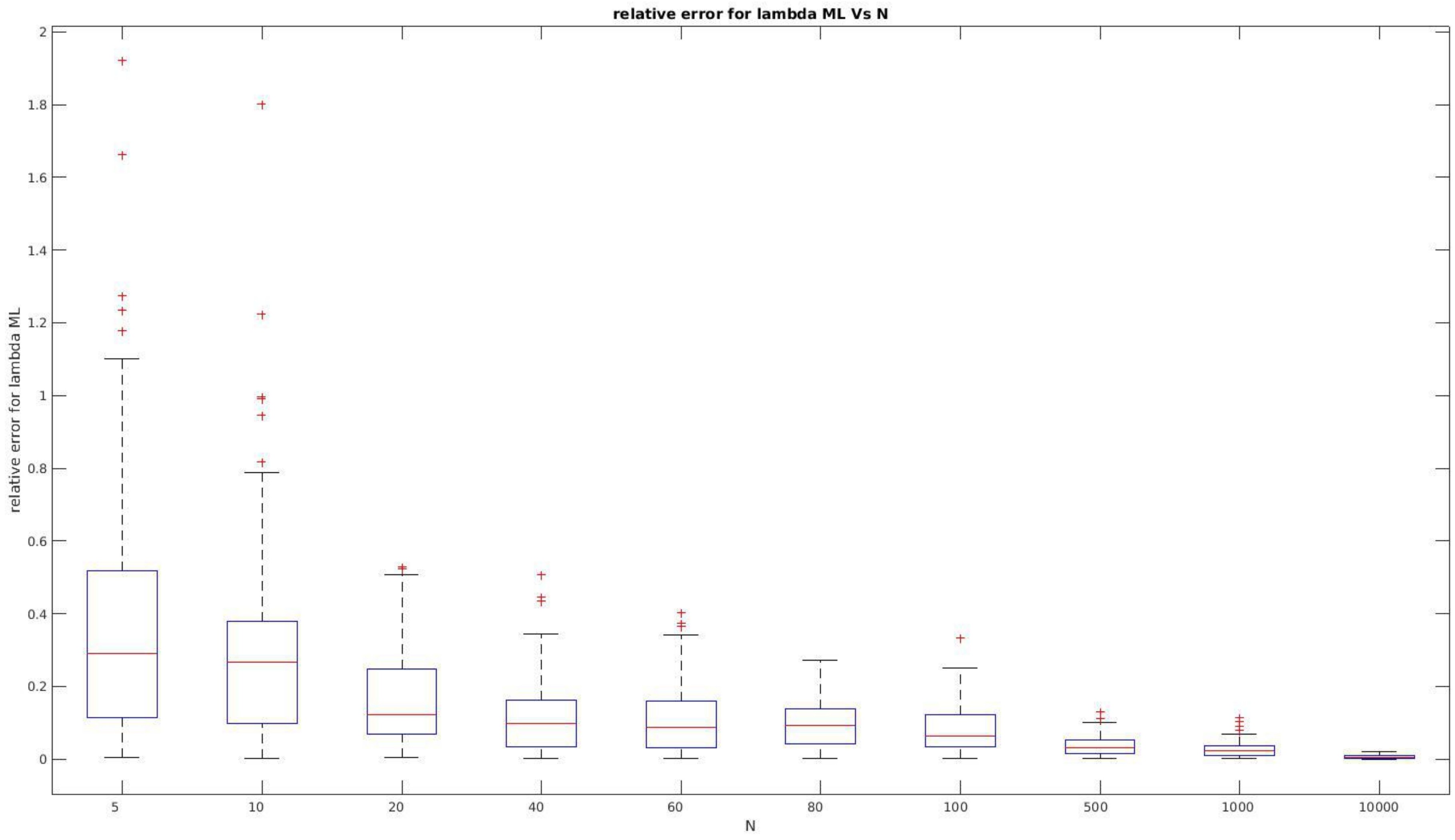
$$\Rightarrow \hat{\lambda}^{\text{posterior Mean}} = \int_0^{\infty} \frac{\lambda^{n+\alpha-1} \cdot e^{-\lambda(\beta + \Sigma y_i)}}{\frac{(n+\alpha-1)!}{(\beta + \Sigma y_i)^{n+\alpha}}} \cdot \lambda \, d\lambda$$

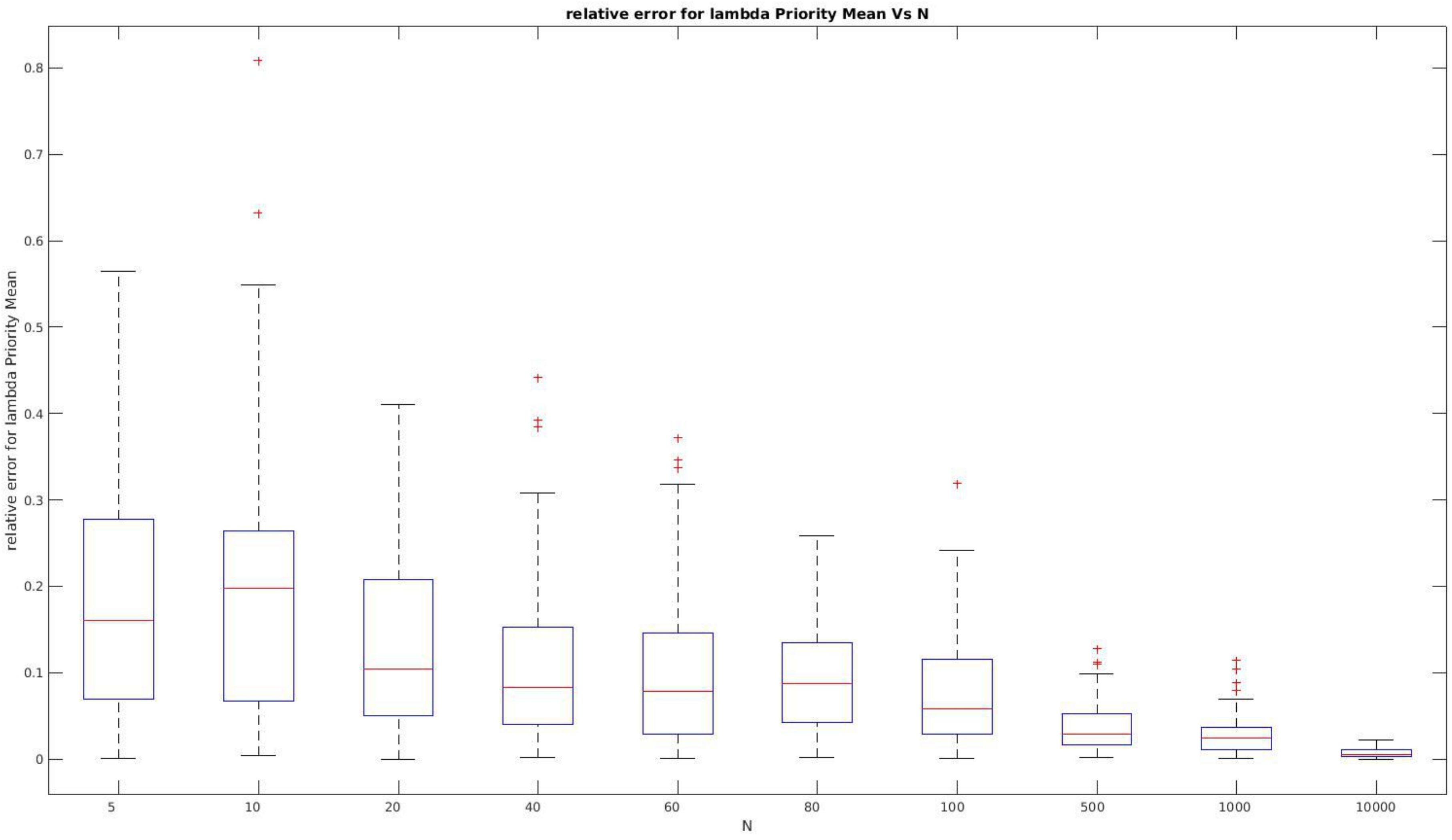
$$\Rightarrow \hat{\lambda}^{\text{posterior Mean}} = \frac{(\beta + \Sigma y_i)^{n+\alpha}}{(n+\alpha-1)!} \int_0^{\infty} \lambda^{n+\alpha} \cdot e^{-\lambda(\beta + \Sigma y_i)} \cdot d\lambda$$

$\Rightarrow \quad \hat{\lambda}^{\text{posterior Mean}} = \dfrac{(\beta + \Sigma y_i)^{n+\alpha}}{(n+\alpha-1)!} \left(\dfrac{(n+\alpha)!}{(\beta + \Sigma y_i)^{n+\alpha+1}} \right)$

$\left(\text{as } \int_0^\infty x^n e^{-ax}\, dx = \dfrac{n!}{a^{n+1}} \text{ for } n > 0 \right)$

$\rightarrow \quad \boxed{\hat{\lambda}^{\text{posterior Mean}} = \dfrac{n+\alpha}{\beta + \Sigma y_i}}$


relative error for lambda ML Vs N
relative error for lambda ML
N
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
5
10
20
40
60
80
100
500
1000
10000


relative error for lambda Priority Mean Vs N
relative error for lambda Priority Mean
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
5
10
20
40
60
80
100
500
1000
10000
N

* clearly as 'N' increases the error decreases from the box-plot graphs.

* Among the two estimates, $\hat{\lambda}$ ~~Maximum Prior~~ Posterior Mean is best one as error will be minimum in this case (from box-plot graph).

3. given N random vectors $x_1, x_2 \dots x_N$

where $x_i = \begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} r\cos\theta_i \\ r\sin\theta_i \end{pmatrix}$ $\boxed{0 \le \theta_i < 2\pi}$

$$f_X(x, x_2 \dots x_N) = \pi f_X$$

$$L = \prod_{i=1}^{N} \frac{1}{\sqrt{(2\pi)^2 |\Sigma|}} e^{-(x_i - \mu)^T \Sigma^{-1} (x_i - \mu)}$$

$$\log L = -N\log(2\pi) - \frac{N}{2}\log E - \frac{1}{2}\sum_{i=1}^{N} (x_i - \mu)^T \Sigma^{-1} (x_i - \mu)$$

where μ = mean matrix and Σ covariance matrix

for Max Likelihood;

$$\frac{\partial LL}{\partial \mu} = \frac{\partial LL}{\partial \Sigma} = 0$$

$$\frac{\partial LL}{\partial \mu} = -\sum_{i=1}^{N} [\Sigma^{-1}(x_i - \mu)] = 0$$

$$\Sigma x_i - N\mu = 0$$

$$\boxed{\hat{\mu} = \frac{\Sigma x_i}{N} = \bar{x}}$$

we have

$$\frac{\partial x^t A x}{\partial A} = \frac{\partial}{\partial A} tr[x^T x A]$$

since $x^t A x$ is scalar and trace is invariant under Cyclic permutations

$$= x x^T$$

$$\log L = -\frac{N}{2}\log|\Sigma| - \frac{1}{2}\sum_{i=1}^{N} (x_i - \mu)^T \Sigma^{-1} (x_i - \mu)$$

Since $\frac{\partial \log|\Sigma|}{\partial E} = \Sigma^{-T}$

$$\log L = C + \frac{N}{2}\log|\Sigma^{-1}| - \frac{1}{2}\sum_{i=1}^{N} tr[x_i - \mu][x_i - \mu]^T \Sigma^{-1}$$

$$\frac{\partial \log L}{\partial \Sigma^{-1}} = N\Sigma - \sum_{i=1}^{N} (x_i - \hat{\mu})(x_i - \hat{\mu})^T = 0$$

$$\hat{\Sigma} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{\mu})(x_i - \hat{\mu})^T$$ (N is very large)

(i) $x_i = \begin{bmatrix} r\cos\theta_i \\ r\sin\theta_i \end{bmatrix}$ $\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$

$$\mu_1 = E(r\cos\theta) = \int_0^{2\pi} r\cos\theta \frac{d\theta}{2\pi} = \frac{r}{2\pi}[\sin\theta]_0^{2\pi} = 0$$

$$\mu_2 = E(r\sin\theta) = \int_0^{2\pi} r\sin\theta \frac{d\theta}{2\pi} = \frac{r}{2\pi}[-\cos\theta]_0^{2\pi} = 0$$

(or) directly $N \to \infty$ in mean

(ii) $C = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}$

$$\sigma_1^2 = \int_0^{2\pi}(x-\mu)^2\frac{d\theta}{2\pi} = \int_0^{2\pi}(r\cos\theta)^2\frac{d\theta}{2\pi} = \frac{r^2}{2\pi}\int_0^{2\pi}\cos^2\theta = \frac{r^2}{2}$$

$$\sigma_2^2 = \int_0^{2\pi}(y-\mu)\frac{d\theta}{2\pi} = \int_0^{2\pi}(r\sin\theta)^2\frac{d\theta}{2\pi} = \frac{r^2}{2\pi}\int_0^{2\pi}\sin^2\theta = \frac{r^2}{2}$$

$$\sigma_{12} = E([x-E(x)][y-E(y)]) = E[r\cos\theta\, r\sin\theta] = \int_0^{2\pi}\frac{r^2}{2\pi}\sin\theta\cos\theta = 0$$

(or) from derived formula

$$\hat{\Sigma} = \frac{1}{N}\begin{bmatrix} \sum_{i=1}^{N} r^2\cos^2\theta_i & r^2\cos\theta_i\sin\theta_i \\ r^2\sin\theta_i\cos\theta_i & r^2\sin^2\theta_i \end{bmatrix} \quad \lim_{N\to\infty} \frac{\sum\cos^2\theta_i}{N} = \frac{\sum\sin^2\theta_i}{N} = \frac{1}{2}$$

$$C = \frac{r^2}{2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$p(x;\mu,C) = \frac{1}{\sqrt{(\sqrt{2}\pi\left(\frac{r^2}{2}\right))^2}}\exp\left[\frac{-1}{2}(x^T)(x)\left(\frac{2}{r^2}\right)\right]$$

$$= \frac{2}{\pi r^4}\exp\left(-\frac{(x^2+y^2)}{r^2}\right) = \frac{2}{\pi r^4}e^{-1}$$

$\underline{3b}$ $\rightarrow$ Mode of the gaussian with R^2 is situated at origin (i.e.) $(x=0, y=0)$ since x^2+y^2 is minimum of origin.

No, the gaussian doesn't fit the data. It's a bad model as we are trying for fit points into a gaussian distribution. Mode $\begin{pmatrix}0\\0\end{pmatrix}$ is not in our data sample.

Problem - 4 :-

(a) Given, Random Variable X has uniform distribution over $(0, \theta)$

So, $P(x_i|\theta) = \begin{cases} \frac{1}{\theta} & 0 \le x_i \le \theta \\ 0 & \text{otherwise} \end{cases}$

$\therefore P(x_1, x_2 \ldots x_N|\theta) = \frac{1}{\theta} \cdot \frac{1}{\theta} \cdots \frac{1}{\theta} = \frac{1}{\theta^N}$

'P' is maximised by choosing θ to be as small as possible but θ clearly must be at least as large as the the largest observed value of $x_1, x_2 \ldots x_N$

So, $\theta = \max\{x_1, x_2 \ldots x_N\}$

Let $\theta' = \max(x_1, x_2, \ldots x_N)$

$\Rightarrow \theta = \theta'$

$\therefore$ Maximum - likelihood estimated $\hat{\theta}^{ML}$ is θ'

Also given, $P(\theta) \propto \left(\frac{\theta_m}{\theta}\right)^\alpha$ for $\theta \ge \theta_m$ and

$P(\theta) = 0$, otherwise

$\Rightarrow P(\theta) = \begin{cases} K\left(\frac{\theta_m}{\theta}\right)^\alpha & \theta \ge \theta_m \\ 0 & \text{otherwise} \end{cases}$

For $\hat{\theta}^{MAP}$,

$$P(\theta|\text{Data}) = \frac{P(\text{Data}|\theta) \cdot P(\theta)}{P(\text{Data})}$$

$$\therefore P(\theta|x_1, x_2 \ldots x_n) = \frac{P(x_1, x_2 \ldots x_n|\theta) \cdot P(\theta)}{\int_{-\infty}^{\infty} P(x_1, x_2 \ldots x_n|\theta) \cdot P(\theta)\, d\theta}$$

$$\Rightarrow P(\theta|\{x_i\}_1^n) = \frac{\left(\frac{1}{\theta^N}\right) \cdot K\left(\frac{\theta_m}{\theta}\right)^\alpha}{\int_{-\infty}^{\infty} \left(\frac{1}{\theta^N}\right) \cdot K\left(\frac{\theta_m}{\theta}\right)^\alpha d\theta} \rightarrow \theta \ge \theta_m$$

$\theta = \theta'$

$$= \frac{\left(\frac{1}{\theta^N}\right) \cdot \cancel{K}\left(\frac{\cancel{\theta_m}}{\theta}\right)^\alpha}{\int_{\max(\theta', \theta_m)}^{\infty} \left(\frac{1}{\theta^N}\right) \cdot \cancel{K}\left(\frac{\cancel{\theta_m}}{\theta}\right)^\alpha d\theta} = \frac{\frac{1}{\theta^{N+\alpha}}}{\int_{\max(\theta', \theta_m)}^{\infty} \frac{1}{\theta^{N+\alpha}}\, d\theta}$$

$$= \frac{\left(\frac{1}{\theta^{N+\alpha}}\right)}{\left[\frac{1}{1-(N+\alpha)} \cdot \frac{1}{\theta^{N+\alpha-1}}\right]_{\max(\theta', \theta_m)}^{\infty}} = \frac{(1-N-\alpha)\left[\max(\theta', \theta_m)\right]^{N}}{\theta^{N+\alpha}}$$

$$P(\theta | \{x_i\}_{i=1}^{n}) = \frac{[1-(N+\alpha)]\left[\max(\theta', \theta_m)\right]^{N+\alpha-1}}{\theta^{N+\alpha}}$$ posterior

$\hat{\theta}^{MAP}$ should maximize the posterior distribution

For $P(\theta | \{x_i\}_1^n)$ to be maximum θ should be minimum also

$\theta \geq \{\max(\theta', \theta_m)\}$ $\therefore \hat{\theta}^{MAP} = \max(\theta', \theta_m)$

(b) As the sample size tends to infinity, $\hat{\theta}^{MAP}$ may or may not be equal to $\hat{\theta}^{ML}$ as θ' may or may not be greater to θ_m. It is not desirable.

(c) Now, $\hat{\theta}^{\text{posterior Mean}} = \int_{\max(\theta', \theta_m)}^{\infty} \theta \cdot \frac{(1-(N+\alpha))(\max(\theta', \theta_m))^{N+\alpha-1}}{\theta^{N+\alpha}} \, d\theta$

Let $c = (1-(N+\alpha))(\max(\theta', \theta_m))^{N+\alpha-1}$

$$\Rightarrow \hat{\theta}^{\text{posterior Mean}} = \int_{\max(\theta', \theta_m)}^{\infty} \frac{c}{\theta^{N+\alpha-1}} \, d\theta$$

$$= c \int_{\max(\theta', \theta_m)}^{\infty} \frac{1}{\theta^{N+\alpha-1}} \, d\theta = c \cdot \frac{1}{2-\alpha-N} \cdot \frac{1}{\theta^{N+\alpha-2}} \Bigg]_{\max(\theta', \theta_m)}^{\infty}$$

$$= \cancel{1-(N+\alpha)} \frac{c}{2-(N+\alpha)} \cdot \frac{1}{(\max(\theta', \theta_m))^{N+\alpha-2}}$$

$$= \frac{1-(N+\alpha)}{2-(N+\alpha)} \cdot \frac{(\max(\theta', \theta_m))^{N+\alpha-1}}{(\max(\theta', \theta_m))^{N+\alpha-2}}$$

$$\therefore \hat{\theta}^{\text{posterior mean}} = \frac{1-(N+\alpha)}{2-(N+\alpha)} \cdot (\max(\theta', \theta_m))$$

(d) As, $N \to \infty$ $\hat{\theta}^{\text{Posterior Mean}} \to \max(\theta', \theta_m)$. This may not or may be equal to $\hat{\theta}^{ML}$ (as θ' may or may not be greater than θ_m). So, It is not desirable.